7-2014

Is your Ontology as Hard as you Think? Rewriting Ontologies into Simpler DLs

David Carral

Cristina Feier

Ana Armas Romero

Bernardo Cuenca-Grau

Pascal Hitzler
pascal.hitzler@wright.edu

See next page for additional authors

Authors
David Carral, Cristina Feier, Ana Armas Romero, Bernardo Cuenca-Grau, Pascal Hitzler, and Ian Horrocks

Is Your Ontology as Hard as You Think? Rewriting Ontologies into Simpler DLs *

David Carral[1], Cristina Feier[2], Ana Armas Romero[2], Bernardo Cuenca Grau[2], Pascal Hitzler[1], and Ian Horrocks[2]

[1] Department of Computer Science, Wright State University, Dayton US
[2] Department of Computer Science, University of Oxford, Oxford UK

Abstract. We investigate cases where an ontology expressed in a seemingly hard DL can be polynomially reduced to one in a simpler logic, while preserving reasoning outcomes for classification and fact entailment. Our transformations target the elimination of inverse roles, universal and existential restrictions, and in the best case allow us to rewrite the given ontology into one of the OWL 2 profiles. Even if an ontology cannot be fully rewritten into a profile, in many cases our transformations allow us to exploit further optimisation techniques. Moreover, the elimination of some out-of-profile axioms can improve the performance of modular reasoners, such as MORe. We have tested our techniques on both classification and data reasoning tasks with encouraging results.

1 Introduction

State-of-the-art DL reasoners such as Pellet [20], JFact, FaCT^{++} [23], RacerPro [10], and HermiT [16] are highly-optimised for classification and have been exploited successfully in many applications. In a recent evaluation campaign, these reasoners exhibited excellent performance on a corpus with over $1,000$ ontologies, as they were able to classify 75%-85% of the corpus in less than 10 seconds when running on stock hardware [9,3].

However, notwithstanding extensive research into optimisation techniques, DL reasoning remains a challenge in practice. Indeed, the aforementioned evaluation also revealed that many ontologies are still hard for reasoners to classify. Furthermore, due to the high worst-case complexity of reasoning, systems are inherently not robust, and even minor changes to ontologies can have a significant effect on performance. Finally, the limitations of DL reasoners become even more apparent when reasoning with ontologies and large datasets.

These issues have motivated a growing interest in lightweight DLs: weaker logics that enjoy more favourable computational properties. Among these are the OWL 2 profiles [15]. Standard reasoning tasks, such as classification and fact entailment, are feasible in polynomial time for all profiles, and a number of highly scalable reasoners have been developed [24,12,18,2,4]. Unfortunately,

*This paper extends our results in [6]. It is accompanied by a technical report which contains all proofs (available at: http://www.cs.ox.ac.uk/isg/TR/TRsafeshoiqDL.pdf) many ontologies fall outside the OWL 2 profiles, and we are forced to resort to a fully-fledged reasoner if a completeness guarantee is required.

In this paper, we propose techniques to (at least partially) rewrite ontologies in the direction of the OWL 2 profiles, specifically EL and RL. All rewritings are polynomial and preserve classification and fact entailment.

In Section 3, we consider rewritings that are applicable to $\mathcal{SHOIQ}$ [11] and that can transform non-EL axioms into EL by elimination of inverse roles and universal restrictions. If all non-EL axioms can be rewritten, we can provide completeness guarantees using only an EL reasoner. Otherwise, the rewritings can still improve the performance of OWL reasoners by enabling the use of optimisations applicable only in the absence of certain constructs and/or the effectiveness of modular reasoners such as MORe [1].

In Section 4, we focus on Horn ontologies and consider rewritings into OWL 2 RL. The RL profile is tightly connected to Datalog, and hence existential restrictions $\exists R.C$ occurring positively in axioms are disallowed, unless C is a singleton nominal $\{o\}$. We show that when R fulfills certain conditions, such concepts $\exists R.C$ can be rewritten into existential restrictions over nominals as accepted in OWL 2 RL; we call such roles R *reuse-safe*. In the limit case where all roles are reuse-safe, the ontology can be polynomially rewritten into RL; if, additionally, the ontology contains no cardinality constraints, it can also be rewritten into EL. Furthermore, if only some roles are reuse-safe, they can be treated by (hyper-)tableau reasoners in an optimised way, potentially reducing the size of the constructed pre-models and improving reasoning times.

Our experiments over a large ontology repository reveal that our techniques can lead to substantial improvements in classification times for both standard and modular reasoners. Furthermore, we show that many ontologies contain only reuse-safe roles and hence can be fully rewritten into RL; thus, highly scalable RL triple stores can be exploited for large-scale data reasoning.

2 Preliminaries

A *signature* consists of disjoint countable sets of *individuals* N_{I}, *atomic concepts* N_{C} and *atomic roles* N_{R}. A *role* is an element of $N_{\mathsf{R}} \cup \{R^- | R \in N_{\mathsf{R}}\}$. The function $\mathsf{Inv}(\cdot)$ is defined over roles as follows, where $R \in N_{\mathsf{R}}$: $\mathsf{Inv}(R) = R^-$ and $\mathsf{Inv}(R^-) = R$. An *RBox* $\mathcal{R}$ is a finite set of *RIAs* $R \sqsubseteq R'$ and *transitivity axioms* $\mathsf{Tra}(R)$, with R and R' roles. We denote with $\sqsubseteq_{\mathcal{R}}$ the minimal relation over roles in $\mathcal{R}$ s.t. $R \sqsubseteq_{\mathcal{R}} S$ and $\mathsf{Inv}(R) \sqsubseteq_{\mathcal{R}} \mathsf{Inv}(S)$ hold if $R \sqsubseteq S \in \mathcal{R}$. We define $\sqsubseteq^*_{\mathcal{R}}$ as the reflexive-transitive closure of $\sqsubseteq_{\mathcal{R}}$. A role R is *transitive* in $\mathcal{R}$ if there is a role S such that $S \sqsubseteq^*_{\mathcal{R}} R$, $R \sqsubseteq^*_{\mathcal{R}} S$ and either $\mathsf{Tra}(S) \in \mathcal{R}$ or $\mathsf{Tra}(\mathsf{Inv}(S)) \in \mathcal{R}$. A role R is *simple* in $\mathcal{R}$ if no transitive role S exists s.t. $S \sqsubseteq^*_{\mathcal{R}} R$. The set of $\mathcal{SHOIQ}$ *concepts* is the smallest set containing $A \in N_{\mathsf{C}}$, $\top$, $\bot$, $\{o\}$ (nominal), $\neg C$ (negation), $C \sqcap D$ (conjunction), $C \sqcup D$ (disjunction), $\exists R.C$ (existential restriction), $\forall R.C$ (universal restriction), $\leqslant nS.C$ (at-most restriction), and $\geqslant nR.C$ (at-least restriction), for $A \in N_{\mathsf{C}}$, C and D $\mathcal{SHOIQ}$ concepts, $o \in N_{\mathsf{I}}$, R a role and S a simple role, and $n \geqslant 0$. A *literal concept* is either atomic or the negation of an atomic concept. A *TBox* $\mathcal{T}$ is a finite set of GCIs $C \sqsubseteq D$ with C, D concepts. An *ABox* $\mathcal{A}$ is a finite set of assertions $C(a)$ (concept assertion), $R(a, b)$ (role assertion), $a \approx b$ (equality assertion), and $a \not\approx b$ (inequality assertion), with C a concept, R a role and a, b individuals. A *fact* is either a concept assertion $A(a)$ with A atomic, a role assertion, an equality assertion, or an inequality assertion. An ontology is a triple $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$. The semantics is standard [11].

We assume familiarity with standard conventions for naming DLs, and we just provide here a definition of the OWL 2 profiles. A $\mathcal{SHOIQ}$ ontology is:

- EL if *(i)* it does not contain inverse roles, negation (other than $\bot$), disjunction, at-most restrictions and at-least restrictions; and *(ii)* every universal restriction appears only in a GCI of the form $\top \sqsubseteq \forall R.C$.
- RL if each GCI $C \sqsubseteq D$ satisfies *(i)* C does not contain negation as well as universal, at-least, and at-most restrictions; *(ii)* D does not contain negation (other than $\bot$), union, existential restrictions (other than of the form $\exists R.\{o\}$), at-least restrictions, and at-most restrictions with $n > 1$.
- QL if it does not contain transitivity and for each GCI $C \sqsubseteq D$ *(i)* C is either atomic or $\exists R.\top$; *(ii)* D is of the form $\bigsqcap_{i=1}^{n} B_i$ with each B_i either a literal concept, or $\bot$, or of the form $\exists R.A$ with R a role and A either atomic or $\top$.

Classification of $\mathcal{O}$ is the task of computing all subsumptions $\mathcal{O} \models A \sqsubseteq B$ with $A \in N_{\mathsf{C}} \cup \{\top\}$, and $B \in N_{\mathsf{C}} \cup \{\bot\}$. Fact entailment is to check whether $\mathcal{O} \models \alpha$, for α a fact. Both problems are reducible to ontology unsatisfiability.

3 Rewriting Ontologies into OWL 2 EL

In this section, we propose techniques for transforming non-EL axioms into EL. Whenever possible, inverse roles are replaced with fresh symbols and the ontology is extended with axioms simulating their possible effects. At the same time, we attempt to transform positive occurrences of universal restrictions into negative occurrences of existential restrictions while inverting the relevant role.

Preprocessing Our first step is to bring $\mathcal{O}$ into a suitable normal form. Normalisation facilitates further rewriting steps, and allows us to identify syntactically non-EL axioms which have a direct correspondence in EL.

Definition 1. *A normalised GCI is of the form:* $\bigsqcap_{i=1}^{n} C_i \sqsubseteq \bigsqcup_{j=1}^{m} D_j$*, where each* C_i *is one of* $\top$*,* A*, or* $\exists R.A$*, and each* D_j *is one of* $\bot$*,* A*,* $\{o\}$*,* $\exists R.A$*,* $\exists R.\{o\}$*,* $\forall R.A$*, or* $\leqslant kR.B$*, with* A *(*B*) an atomic (literal) concept,* R *a role,* $o \in N_I$*, and* $k \geqslant 1$*; each concept* C_i *(resp.* D_j*) is said to occur positively (negatively) in the GCI. An ontology* $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ *is normalised if* $\mathcal{A}$ *only contains facts and each GCI in* $\mathcal{T}$ *is normalised. Furthermore,* $\mathcal{O}$ *is* Horn *if* $m = 1$ *for each GCI in* $\mathcal{O}$ *and each at-most restriction* $\leqslant kR.C$ *satisfies* $k = 1$*.*

Proposition 1. *There exists a polynomial transformation* Υ *over* $\mathcal{SHOIQ}$ *ontologies* $\mathcal{O}$ *such that:* (i) $\Upsilon(\mathcal{O})$ *is normalised;* (ii) $\mathcal{O}$ *is satisfiable iff* $\Upsilon(\mathcal{O})$ *is satisfiable; and* (iii) *if* $\mathcal{O}$ *is EL (resp. RL, or QL), then so is* $\Upsilon(\mathcal{O})$*.*[3]

[3]Under a trivial relaxation of QL syntax. Please see the technical report for details.

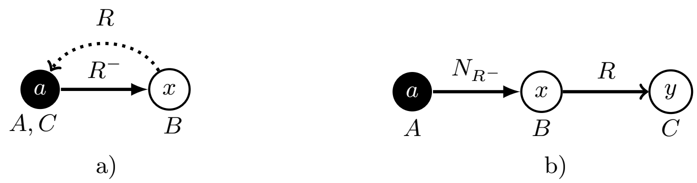

Fig. 1. When rewriting away inverse roles leads to missing entailments

Inverse Rewritability Satisfiability of $\mathcal{SHOIQ}$ ontologies is NEXPTIME-complete, whereas for $\mathcal{SHOQ}$ it is EXPTIME-complete; thus, in general, inverse roles cannot be faithfully eliminated from $\mathcal{SHOIQ}$ ontologies by a polynomial transformation. The following example illustrates that an obstacle to rewritability is the interaction between inverses and at-most restrictions.

Example 1. Consider $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$, with $\mathcal{R} = \emptyset$, $\mathcal{A} = \{A(a)\}$, and $\mathcal{T}$ as follows:

$$\mathcal{T} = \{A \sqsubseteq \exists R^{-}.B; \quad B \sqsubseteq \exists R.C; \quad B \sqsubseteq \leqslant 1\, R.\top\}$$

Note that $\mathcal{O} \models C(a)$. In every model $(\Delta_{\mathcal{I}}, \cdot^{\mathcal{I}})$, object $a^{\mathcal{I}}$ must be R^{-}-connected to some $x \in B^{\mathcal{I}}$ (due to the first axiom in $\mathcal{T}$); also, x must be R-connected to some $y \in C^{\mathcal{I}}$ (due to the second axiom). Then, for the last axiom to be satisfied, $a^{\mathcal{I}}$ and y must be identical; thus, $a^{\mathcal{I}} \in C^{\mathcal{I}}$. Figure 1 a) depicts such a model. Consider now $\mathcal{O}'$ obtained from $\mathcal{O}$ by replacing R^{-} with a fresh atomic role $N_{R^{-}}$. Then, $\mathcal{O}' \not\models C(a)$, and Fig. 1 b) depicts a model of $\mathcal{O}'$ not satisfying $C(a)$. Extending $\mathcal{O}'$ with EL axioms to simulate the interaction between inverses and cardinality restrictions (and recover the missing entailment) seems infeasible. ♢

We next propose sufficient conditions for inverse roles to be rewritable.

Definition 2. *Let $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ be a normalised $\mathcal{SHOIQ}$ ontology. A (possibly inverse) role R is* generating *in $\mathcal{O}$, if an existential restriction over a role $R' \sqsubseteq^{*}_{\mathcal{R}} R$ occurs positively in a GCI.*

An inverse role S^{-} in $\mathcal{O}$ is rewritable *if for each $X \in \{S, S^{-}\}$ occurring in at-most restrictions in $\mathcal{O}$ we have that $\mathsf{Inv}(X)$ is not generating in $\mathcal{O}$.*

Intuitively, roles in positive occurrences of existential restrictions are those "inducing" the edges between individuals and their successors in a canonical forest model; a role R is generating if it is a super-role of one such R'. Our condition ensures that "backwards" edges in such a canonical model of $\mathcal{O}$ (i.e., those induced by an inverse role) cannot invalidate an at-most cardinality restriction. As we will show later on (c.f. Theorems 1 and 2), when all inverse roles in a $\mathcal{SHOIQ}$ ontology are rewritable we can faithfully rewrite the ontology into $\mathcal{SHOQ}$ by means of a polynomial transformation.

$$
\begin{aligned}
C \sqsubseteq D \sqcup \forall R.A &\Rightarrow \{C \sqsubseteq D \sqcup X,\ \exists \mathsf{Inv}(R).X \sqsubseteq A\} \\
&\quad \text{if } R \text{ is not generating} \\
C \sqsubseteq D \sqcup \forall R.A &\Rightarrow \{C \sqsubseteq D \sqcup X,\ \exists \mathsf{Inv}(R).X \sqsubseteq A,\ X \sqsubseteq \forall R.A\} \\
&\quad \text{if } \mathsf{Inv}(R) \text{ is generating and } R \text{ is generating} \\
C \sqcap \exists R.A \sqsubseteq D &\Rightarrow \{C \sqcap X \sqsubseteq D, A \sqsubseteq \forall \mathsf{Inv}(R).X\} \\
&\quad \text{if } \mathsf{Inv}(R) \text{ is generating and } R \text{ is not generating} \\
C \sqcap \exists R.A \sqsubseteq D &\Rightarrow \{C \sqcap X \sqsubseteq D,\ A \sqsubseteq \forall \mathsf{Inv}(R).X,\ \exists R.A \sqsubseteq X\} \\
&\quad \text{if } \mathsf{Inv}(R) \text{ is generating and } R \text{ is generating}
\end{aligned}
$$

Fig. 2. Rewrite rules. X is fresh in each rule application.

The Transformation Before presenting our transformation formally, we exemplify how universal restrictions can be replaced with (negative occurrences of) existential restrictions if the relevant roles are not generating.

Example 2. Consider $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ where $\mathcal{R} = \{R \sqsubseteq S^-\}$, $\mathcal{A} = \{A(a); S(a,b)\}$, and $\mathcal{T} = \{A \sqsubseteq \forall S.B;\ \ B \sqsubseteq \exists R.C;\ \ \exists S.B \sqsubseteq D;\ \ C \sqcap D \sqsubseteq \bot\}$.

Clearly, $\mathcal{O}$ is unsatisfiable. Furthermore, it does not contain at most restrictions, and hence S^- is rewritable. We first extend $\mathcal{O}$ with logically redundant axioms, which make explicit information that may be lost when replacing inverses with fresh symbols. Thus, we extend $\mathcal{T}$ with $\exists S^-.A \sqsubseteq B$, and $B \sqsubseteq \forall S^-.D$; furthermore, we extend $\mathcal{R}$ with $R^- \sqsubseteq S$; and finally, $\mathcal{A}$ with the assertion $S^-(b,a)$.

An important observation is that S is not generating. As a result, we can dispense with axiom $A \sqsubseteq \forall S.B$. Then we replace S^- with a fresh symbol N_{S^-} and R^- with N_{R^-}. The resulting $\mathcal{O}' = (\mathcal{R}', \mathcal{T}', \mathcal{A}')$ is as follows:

$$
\begin{aligned}
\mathcal{R}' &= \{R \sqsubseteq N_{S^-};\ \ N_{R^-} \sqsubseteq S\} \\
\mathcal{T}' &= \{\exists N_{S^-}.A \sqsubseteq B;\ \ B \sqsubseteq \exists R.C;\ \ \exists S.B \sqsubseteq D;\ \ B \sqsubseteq \forall N_{S^-}.D;\ \ C \sqcap D \sqsubseteq \bot\} \\
\mathcal{A}' &= \{A(a);\ \ S(a,b);\ \ N_{S^-}(b,a)\}
\end{aligned}
$$

$\mathcal{O}'$ is unsatisfiable; furthermore it is in EL except for axiom $B \sqsubseteq \forall N_{S^-}.D$. This axiom cannot be dispensed with since S^- is generating, and hence it is needed to propagate information along N_{S^-}-edges in a canonical model. ♢

We next present our transformation. For simplicity, we restrict ourselves to $\mathcal{ALCHOIQ}$ ontologies; later on, we discuss issues associated with transitivity axioms and show how our techniques extend to $\mathcal{SHOIQ}$.

Definition 3. *Let $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ be a normalised $\mathcal{ALCHOIQ}$ ontology. The ontology $\Xi(\mathcal{O}) = (\mathcal{R}', \mathcal{T}', \mathcal{A}')$ is obtained as follows:*

1. Axiom Rewriting*: the ontology $\mathcal{O}_e = (\mathcal{R}_e, \mathcal{T}_e, \mathcal{A}_e)$ is defined as follows:*
   - *$\mathcal{R}_e$ extends $\mathcal{R}$ with an axiom $\mathsf{Inv}(R) \sqsubseteq \mathsf{Inv}(S)$ for each $R \sqsubseteq S$ in $\mathcal{R}$;*
   - *$\mathcal{T}_e$ is obtained from $\mathcal{T}$ by applying exhaustively the rewrite rules in Fig. 2, and deleting all axioms mentioning a non-generating role R for which $\mathcal{A}$ contains no assertion $S(a,b)$ with $S \sqsubseteq R$ or $S \sqsubseteq \mathsf{Inv}(R)$;*

– $\mathcal{A}_e$ *extends* $\mathcal{A}$ *with an assertion* $R^-(b,a)$ *for each* $R(a,b) \in \mathcal{A}$.

2. Inverse Replacement: $\Xi(\mathcal{O}) = (\mathcal{R}', \mathcal{T}', \mathcal{A}')$ *is obtained from* $\mathcal{O}_e$ *by replacing each occurrence of an inverse role that is rewritable in* $\mathcal{O}_e$ *with a fresh role.*

The first step in the transformation extends the ontology with axioms that simulate the effect of inverse roles, which are eliminated in the second step. Furthermore, the rewrite rules in Fig. 2 are designed to eliminate "harmless" occurrences of universal restrictions (see Example 2).

Theorem 1. *Let* $\mathcal{O}' = \Xi(\mathcal{O})$*. Then,* $\mathcal{O}'$ *is of size polynomial in the size of* $\mathcal{O}$ *and it is satisfiable iff* $\mathcal{O}$ *is satisfiable. Furthermore, if* $\mathcal{O}$ *contains only rewritable inverse roles, then* $\mathcal{O}'$ *is an* $\mathcal{ALCHOQ}$ *ontology. Finally, if* $\mathcal{O}$ *is Horn and it satisfies the following properties, then* $\mathcal{O}'$ *is EL:*

1. *it does not contain at-most restrictions;*
2. *concepts* $\forall R.A$ *with* R *generating occur only in axioms* $\top \sqsubseteq \forall R.A$*; and*
3. *if* $\exists R.A$ *occurs negatively, either* $A = \top$*, or* $\mathsf{Inv}(R)$ *is not generating.*

Theorem 1 identifies a class of $\mathcal{ALCHOIQ}$ ontologies which can be transformed into equisatisfiable $\mathcal{ALCHOQ}$ ontologies and for which standard reasoning is feasible in ExpTime (in contrast to NExpTime). This result can be exploited for optimisation: tableaux reasoners employ pairwise blocking techniques over $\mathcal{ALCHOIQ}$ ontologies, while they rely on more aggressive single blocking for $\mathcal{ALCHOQ}$ inputs, which reduces the size of the constructed pre-models.

The last condition in the theorem establishes sufficient conditions on $\mathcal{O}$ for the transformed ontology $\mathcal{O}'$ to be in EL. A simple case is when $\mathcal{O}$ is in the QL profile of OWL 2, in which case the transformed ontology is guaranteed to be in EL. An interesting consequence of this result is that highly optimised EL reasoners, such as ELK, can be exploited for classifying QL ontologies.

Corollary 1. *If* $\mathcal{O}$ *is a normalised QL ontology, then* $\Xi(\mathcal{O})$ *is in EL.*

In many cases our transformation may only succeed in partially rewriting a ontology into EL (cf. Example 2). Even in these cases, our techniques can have substantial practical benefits (see Evaluation section). As already mentioned, in the absence of inverse roles (hyper-)tableau reasoners may exploit more aggressive blocking techniques. Furthermore, modular reasoning systems such as MORe, which are designed to behave better for ontologies with a large EL subset, benefit from our transformations.

Dealing with Transitivity Axioms The transformation in Definition 3 is not applicable to ontologies with transitivity axioms.

Example 3. Consider $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ with $\mathcal{R} = \{R \sqsubseteq R^-; \mathsf{Tra}(R)\}$, $\mathcal{A} = \{A(a)\}$, and $\mathcal{T} = \{A \sqsubseteq \exists R.B; A \sqsubseteq C; \exists R^-.C \sqsubseteq D\}$. Let $\mathcal{O}' = \Xi(\mathcal{O})$, where we assume that the transitivity axiom $\mathsf{Tra}(R)$ stays unmodified in $\mathcal{O}'$. More precisely, $\mathcal{A}' = \mathcal{A}$, $\mathcal{R}' = \{R \sqsubseteq N_{R^-}; N_{R^-} \sqsubseteq R, Tra(R)\}$, and $\mathcal{T}' = \{A \sqsubseteq \exists R.B; A \sqsubseteq C; \exists N_{R^-}.C \sqsubseteq D; C \sqsubseteq \forall R.D\}$. Then, $\mathcal{O} \models D(a)$, but $\mathcal{O}' \not\models D(a)$; An attempt to recover this entailment by making N_{R^-} transitive does not solve the problem. $\diamondsuit$

To address this issue, we eliminate transitivity before applying our transformation in Definition 3. We consider that ontologies are further normalised s.t. each GCI has at most one negative occurrence of an existential restriction or one positive occurrence of a universal restriction over a non-simple role. We say that such an ontology is *transitivity-normalised*.

Definition 4. *Let $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ be a transitivity normalised $\mathcal{SHOIQ}$ ontology.Let $\Omega(\mathcal{O}) = (\mathcal{R}', \mathcal{T}', \mathcal{A})$ be an ontology in which $\mathcal{R}'$ is obtained from $\mathcal{R}$ by removing all transitivity axioms and $\mathcal{T}'$ is obtained from $\mathcal{T}$ by adding:*

- *for each axiom $C \sqsubseteq D \sqcup \forall R.A$ in $\mathcal{T}$, with R non-simple, and each transitive sub-role S of R: $C \sqsubseteq D \sqcup \forall S.Y_A^S$, $Y_A^S \sqsubseteq \forall S.Y_A^S$, and $Y_A^S \sqsubseteq A$, where Y_A^S is a fresh atomic concept uniquely associated to S and A;*
- *for each axiom $C \sqcap \exists R.A \sqsubseteq D$ in $\mathcal{T}$, with R non-simple, and each transitive sub-role S of R in $\mathcal{R}$: $A \sqsubseteq Z_A^S$, $\exists S.Z_A^S \sqsubseteq Z_A^S$, and $C \sqcap \exists S.Z_A^S \sqsubseteq D$, where Z_A^S is a fresh atomic concept uniquely associated to S and A.*

Lemma 1 establishes the properties of transitivity elimination, and Theorem 2 shows that our techniques extend to a $\mathcal{SHOIQ}$ ontology $\mathcal{O}$ by first applying Ω to $\mathcal{O}$ and then Ξ to the resulting ontology.

Lemma 1. *Let $\mathcal{O}$ be a transitivity normalised $\mathcal{SHOIQ}$ ontology. Then:*

1. *$\Omega(\mathcal{O})$ is satisfiable iff $\mathcal{O}$ is satisfiable.*
2. *$\Omega(\mathcal{O})$ is a normalised $\mathcal{ALCHOIQ}$ ontology; furthermore, $\Omega(\mathcal{O})$ is Horn iff $\mathcal{O}$ is Horn.*
3. *$\Omega(\mathcal{O})$ can be computed in time polynomial in the size of $\mathcal{O}$.*
4. *If $\mathcal{O}$ is EL, then so is $\Omega(\mathcal{O})$.*
5. *If an inverse role R^- is rewritable in $\mathcal{O}$, then it is also rewritable in $\Omega(\mathcal{O})$.*

Theorem 2. *Let $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ be a transitivity normalised $\mathcal{SHOIQ}$ ontology, and let $\mathcal{O}' = \Xi(\Omega(\mathcal{O}))$. Then, $\mathcal{O}'$ satisfies all properties in Theorem 1.*

4 Rewriting Horn Ontologies into OWL 2 RL

We now restrict our attention to Horn ontologies, and consider rewritings into RL. The idea is to identify roles R such that each positive occurrence of a concept $\exists R.C$ can be replaced by an existential restriction $\exists R.\{o\}$ over a nominal (which is allowed in RL) and a fact $C(o)$. We call such roles *reuse-safe*. It is well-known that all roles in an EL ontology satisfy this property, and hence it is possible to faithfully rewrite EL ontologies into RL [17,13,22]. Intuitively, for a role R to be reuse-safe it must be the case that any R-edges in a canonical model of the ontology are irrelevant to the satisfaction of non-EL axioms in the ontology.

Example 4. Consider the following ontology $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ where $\mathcal{R} = \emptyset$, $\mathcal{A} = \{A(a)\}$, and $\mathcal{T}$ consists of the following axioms:

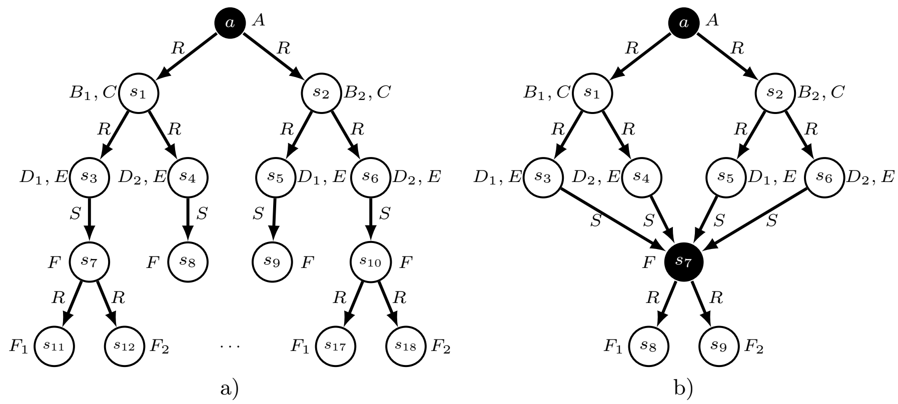

Fig. 3. Decreasing model size by reusing individuals as existential fillers

$$\begin{array}{llllll} A \sqsubseteq \exists R.B_1 & C \sqsubseteq \exists R.D_1 & A \sqsubseteq \forall R.C & E \sqsubseteq \exists S.F & F \sqsubseteq \exists R.F_2 & B_1 \sqcap B_2 \sqsubseteq \bot \\ A \sqsubseteq \exists R.B_2 & C \sqsubseteq \exists R.D_2 & C \sqsubseteq \forall R.E & F \sqsubseteq \exists R.F_1 & F_1 \sqcap F_2 \sqsubseteq \bot & D_1 \sqcap D_2 \sqsubseteq \bot \end{array}$$

Since R is generating and $\mathcal{O}$ has no inverses, we have $\Xi(\mathcal{O}) = \mathcal{O}$. Figure 3 a) depicts a canonical model of $\mathcal{O}$. Role S is reuse-safe since it is not "affected" by non-EL axioms involving universal restrictions. Thus, we can "fold" the model by identifying all nodes with an S-predecessor to a single fresh nominal and obtain a smaller model satisfying the same subsumptions and facts (Fig. 3 b). $\diamondsuit$

Definition 5. *Let $\mathcal{O} = (\mathcal{R}, \mathcal{T}, \mathcal{A})$ be a normalised Horn ontology. A role R in $\mathcal{O}$ is* reuse-safe *if either no existential restriction of the form $\exists R.A$ with $A \in N_C$ occurs positively in $\mathcal{O}$, or each of the following properties hold for each role S:*

- $R \not\sqsubseteq^*_{\mathcal{R}} S$ *and* $R \not\sqsubseteq^*_{\mathcal{R}} \mathsf{Inv}(S)$ *if S occurs in a concept $\leqslant 1\, S.B$;*
- $R \not\sqsubseteq^*_{\mathcal{R}} S$ *if $\mathcal{O}$ contains an axiom $C \sqsubseteq \forall S.B$ with $C \neq \top$;*
- $R \not\sqsubseteq^*_{\mathcal{R}} \mathsf{Inv}(S)$ *if a concept $\exists S.A$ with $A \neq \top$ occurs negatively in $\mathcal{O}$.*

For each concept $\exists R.A$ which occurs positively in $\mathcal{O}$ with R reuse-safe, let $c_{R,A}$ be a fresh individual. Then, $\Psi(\mathcal{O})$ is ontology obtained from $\mathcal{O}$ by replacing each such $\exists R.A$ by $\exists R.\{c_{R,A}\}$ and adding the fact $B(c_{R,A})$ to $\mathcal{A}$.

Theorem 3. *$\Psi(\mathcal{O})$ is satisfiable iff $\mathcal{O}$ is satisfiable, for each $\mathcal{O}$ Horn.*

In practice, system developers can achieve the same goal as our transformation by making their implementations sensitive to reuse-safe roles: to satisfy an axiom involving an existential restriction over such role, a system should reuse a suitable distinguished individual instead of generating a fresh one.

We next analyse the limit case where all roles in a Horn ontology $\mathcal{O}$ are reuse-safe. We show that $\Psi(\mathcal{O})$ is an RL ontology. Furthermore, we can identify a new efficiently-recognisable class of DL ontologies that contains all the OWL 2 profiles, and for which standard reasoning is feasible in polynomial time.

Ontology ID	00047	00461	00463	00470	00484	00533	00660	00760	Fly	CAO
$\mathcal{O}$ (HermiT)	51.41	2.260	t-out	398.38	172.71	39.34	59.02	837.93	937.57	1251.46
$\mathcal{O}'$ (HermiT)	29.93	1.801	422.74	75.55	50.39	20.98	11.62	402.69	10.92	197.5
$\mathcal{O}$ (MORe)	66.55	2.515	t-out	415.17	172.71	5.67	67.26	798.61	954.44	1220.98
$\mathcal{O}'$ (MORe)	29.71	1.310	379.88	2.24	60.94	2.19	12.39	371.68	13.62	186.35

Table 1. Classification times (s) for some ontologies $\mathcal{O}$ and transformed versions $\mathcal{O}'$.

Theorem 4. *For the class $\mathcal{C}$ of Horn ontologies for which all roles are reuse-safe:*

1. *Checking whether a $\mathcal{SHOIQ}$ ontology $\mathcal{O}$ is in $\mathcal{C}$ is feasible in polynomial time;*
2. *Every EL, QL and RL ontology is contained in $\mathcal{C}$;*
3. *$\Psi(\mathcal{O})$ is an RL ontology for each $\mathcal{O} \in \mathcal{C}$; and*
4. *Classification and fact entailment in $\mathcal{C}$ are feasible in polynomial time.*

Finally, it is worth emphasising that, although the transformations Ψ in Definition 3 and Ξ in Sect. 3 are very different and serve rather orthogonal purposes, they are connected in the limit case where all roles are reuse-safe and the ontology does not contain cardinality restrictions.

Proposition 2. *Let $\mathcal{O}$ be a normalised Horn ontology that does not contain at-most restrictions. Then, $\Xi(\Omega(\mathcal{O}))$ is EL iff all roles in $\mathcal{O}$ are reuse-safe.*

5 Evaluation

Classification Experiments We tested all the OWL 2 ontologies in the Oxford Ontology Repository,[4] as well as a "hard" version of the FlyAnatomy ontology, and two additional ontologies from NCBO BioPortal[5]: BIOMODELS and CAO. Several ontologies have a small number of non-$\mathcal{SHOIQ}$ axioms, which we removed for testing. We measured classification times for the latest versions of HermiT (v.1.3.8) and MORe (v.0.1.5) using their standard settings. Experiments were performed on a laptop with 16 GB RAM and Intel Core 2.9 GHz processor running Java v.1.7.0_21, with a timeout of $1h$.

EL rewriting Experiments. Out of the 793 test ontologies, we selected those 70 with inverse roles, and which HermiT takes at least 1s to classify. For each ontology $\mathcal{O}$ we computed a transformed version $\mathcal{O}'$ (see Section 3), and have compared classification times for HermiT and MORe. Out of the 70 test ontologies 50 contained only rewritable inverse roles, which could be successfully eliminated using our transformations, and 10 of these could be fully rewritten into EL. HermiT and MORe exhibited very similar behaviour, and Table 1 presents results for some representative cases. Both reasoners timed out on 8 out of these

[4] http://www.cs.ox.ac.uk/isg/ontologies/
[5] https://bioportal.bioontology.org

50 ontologies, and only on 6 of them after applying our transformations; thus, they succeeded on 2 transformed ontologies that could not be classified in their original form. For 41 of the remaining 42 cases, both reasoners showed either a noticeable improvement or reasoning times very close to the original ones; on average, there was an 3.74 speedup factor for HermiT and 4.37 for MORe. In the case of HermiT, improvements can be explained by the use of a more optimised blocking strategy, which decreased the size of the constructed models by X on average. The improvement in MORe was due, on the one han, to the improvement of HermiT and, on the other hand, to the use of ELK over a larger EL module. Only in one case (BIOMODELS) we observed decreased performance of 21% in HermiT and 78% in MORe. Finally, 20 of the 70 tested ontologies contain non-rewritable inverse roles. As expected, in these cases we obtained no consistent improvement since the presence of inverses forces HermiT to use pairwise blocking; furthermore, in some cases the transformation negatively impacts performance, as it adds a substantial number of axioms to simulate the effect of inverse roles. Hence, it seems that our techniques are clearly beneficial only when all inverse roles are rewritable.

Reuse-safe experiments. From the 793 ontologies in the corpus, we identified 174 Horn ontologies that do not fall within any of the OWL 2 profiles. We have applied our transformation in Definition 3 to these ontologies and found that 53 do not contain unsafe roles and hence are rewritable into RL. Furthermore, in the remaining ontologies 89% of the roles were reuse-safe, on average. We have tested classification times with HermiT over the transformed ontologies, but found that the transformation had a negative impact on performance. This is explained by the introduction of nominals, which forces HermiT to disable *anywhere blocking*. As mentioned in Section 4, it would be more effective to implement safe reuse as a modification of HermiT; this, however, implies non-trivial modifications to the core of the reasoner, which is left for future work.

Data Reasoning Experiments From the 50 ontologies with rewritable inverse roles, we selected those 30 equipped with an ABox, and performed instance retrieval using HermiT on the original and rewritten ontologies. On average, we observed a 3.64 speedup factor. We also tested our rewritings into RL for the LUBM benchmark, which comes with a non-EL ontology that can be fully rewritten into RL. For each dataset, we recorded the times needed to compute the instances of all atomic concepts in the ontology. We compared HermiT over the original ontology and the RL reasoner RDFox[18] over the transformed ontology. HermiT took 3.7s for LUBM(1), and timed out for LUBM(5). In contrast RDFox only required 0.2s for LUBM(1), 1.5s for LUBM(10), and 7.4s for LUBM(20). These results suggest the clear benefits of transforming an ontology to RL and exploiting highly scalable reasoners such as RDFox.

6 Related Work

Several techniques for inverse role elimination in DL ontologies have been developed. Ding et al. [8] propose a polynomial reduction from $\mathcal{ALCI}$ into $\mathcal{ALC}$, which is then extended in [7] to $\mathcal{SHOI}$. Similarly, Song et al. [21] propose a polynomial reduction from $\mathcal{ALCHI}$ to $\mathcal{ALCH}$ KBs to optimise classification. In all of these approaches inverse roles are replaced with fresh symbols and new axioms are introduced to compensate for the loss of implicit inferences. These approaches, however, are not applicable to KBs with cardinality restrictions; furthermore, inverse role elimination heavily relies on the introduction of universal restrictions, and hence they are not well-suited for rewriting into EL. Calvanese et al. [5] propose a transformation from $\mathcal{ALCFI}$ knowledge bases to $\mathcal{ALC}$ which is sound and complete for classification; this technique exhaustively introduces universal restrictions to simulate at-most cardinality restrictions and inverse roles, and hence it is also not targeted towards rewritings into EL; furthermore, this technique is not applicable to KBs with transitive roles or nominals.

Ren at al. proposed technqiues for approximating an OWL ontology into EL [19]; such approximations are, however, incomplete and hence relevant inferences might be lost. Finally, Lutz et al. study rewritability of first-order formulas into EL as a decision problem [14]; the rewritings studied in [14], however, require preservation of logical equivalence, whereas ours preserve satisfiability.

Our techniques in Section 4 extend the so-called combined approach to query answering in EL [13,22]. They are also related to individual reuse optimisations [17], where to satisfy existential restrictions a (hyper-)tableau reasoner tries to reuse an individual from the model constructed thus far. Individual reuse, however, may introduce non-determinism in exchange for a smaller model: if the reuse fails (i.e., a contradiction is derived), the reasoner must backtrack and introduce a fresh individual. In contrast, in the case of reuse-safe roles reuse can be done *deterministically* and hence model size is reduced without the need of backtracking. Finally, Zhou et. al use a very similar transformation as ours to strengthen ontologies and overestimate query answers [25]. It follows from Theorem 4 that the technique in [25] leads to exact answers to atomic queries for Horn ontologies where all roles are reuse-safe.

7 Conclusions and Future Work

We have proposed techniques for rewriting ontologies into the OWL 2 profiles. Our techniques are easily implementable as preprocessing steps in DL reasoners, and can lead to substantial improvements in reasoning times. Furthermore, we have established sufficient conditions for ontologies to be polynomially rewritable into the EL and RL profiles. Thus, for the class of ontologies satisfying our conditions reasoning becomes feasible in polynomial time. There are many avenues to explore for future work. For example, we will investigate extensions of our EL rewriting techniques that are capable of rewriting away disjunctive axioms. Furthermore, we are planning to implement safe reuse in HermiT and evaluate the impact of this optimisation on classification.

Acknowledgements. Work supported by the Royal Society, the EPSRC project Score! and the National Science Foundation under the award TROn: Tractable Reasoning with Ontologies.

References

1. Armas Romero, A., Cuenca Grau, B., Horrocks, I.: MORe: modular combination of OWL reasoners for ontology classification. In: ISWC. pp. 1–16 (2012)
2. Baader, F., Lutz, C., Suntisrivaraporn, B.: CEL - A polynomial-time reasoner for life science ontologies. In: IJCAR. pp. 287–291 (2006)
3. Bail, S., Glimm, B., Gonçalves, R.S., Jiménez-Ruiz, E., Kazakov, Y., Matentzoglu, N., Parsia, B. (eds.): Proc. of the 2nd International Workshop on OWL Reasoner Evaluation (ORE), CEUR, vol. 1015 (2013)
4. Bishop, B., Kiryakov, A., Ognyanoff, D., Peikov, I., Tashev, Z., Velkov, R.: OWLim: A family of scalable semantic repositories. Semantic Web J. 2(1), 33–42 (2011)
5. Calvanese, D., De Giacomo, G., Rosati, R.: A note on encoding inverse roles and functional restrictions in $\mathcal{ALC}$ knowledge bases. In: DL. pp. 69–71. CEUR (1998)
6. Carral, D., Feier, C., Grau, B.C., Hitzler, P., Horrocks, I.: $\mathcal{EL}$-ifying ontologies. In: IJCAR (2014), to appear
7. Ding, Y.: Tableau-based reasoning for Description Logics with inverse roles and number restrictions. Ph.D. thesis, Concordia University, Canada (2008)
8. Ding, Y., Haarslev, V., Wu, J.: A new mapping from $\mathcal{ALCI}$ to $\mathcal{ALC}$. In: Calvanese, D., Franconi, E., Haarslev, V., Lembo, D., Motik, B., Turhan, A., Tessaris, S. (eds.) DL. CEUR Workshop Proceedings, vol. 250 (2007)
9. Gonçalves, R.S., Matentzoglu, N., Parsia, B., Sattler, U.: The empirical robustness of Description Logic classification. In: DL. pp. 197–208 (2013)
10. Haarslev, V., Hidde, K., Möller, R., Wessel, M.: The RacerPro knowledge representation and reasoning system. Semantic Web J. 3(3), 267–277 (2012)
11. Horrocks, I., Sattler, U.: A tableau decision procedure for $\mathcal{SHOIQ}$. J. of Automated Reasoning 39(3), 249–276 (2007)
12. Kazakov, Y., Krötzsch, M., Simancik, F.: Concurrent classification of $\mathcal{EL}$ ontologies. In: ISWC. pp. 305–320 (2011)
13. Kontchakov, R., Lutz, C., Toman, D., Wolter, F., Zakharyaschev, M.: The combined approach to ontology-based data access. In: IJCAI. pp. 2656–2661 (2011)
14. Lutz, C., Piro, R., Wolter, F.: Description Logic TBoxes: Model-theoretic characterizations and rewritability. In: IJCAI. pp. 983–988 (2011)
15. Motik, B., Cuenca Grau, B., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C. (eds.): OWL 2 Web Ontology Language: Profiles. W3C Recommendation (27 October 2009), available at `http://www.w3.org/TR/owl2-profiles/`
16. Motik, B., Shearer, R., Horrocks, I.: Hypertableau reasoning for Description Logics. J. Artificial Intelligence Research (JAIR) 36(1), 165–228 (2009)
17. Motik, B., Horrocks, I.: Individual reuse in Description Logic reasoning. In: IJCAR. pp. 242–258 (2008)
18. Motik, B., Nenov, Y., Piro, R., Horrocks, I., Olteanu, D.: Parallel materialisation of Datalog programs in centralised, main-memory RDF systems. In: AAAI (2014)
19. Ren, Y., Pan, J.Z., Zhao, Y.: Soundness preserving approximation for TBox reasoning. In: AAAI (2010)
20. Sirin, E., Parsia, B., Cuenca Grau, B., Kalyanpur, A., Katz, Y.: Pellet: A practical OWL-DL reasoner. J. Web Semantics (JWS) 5(2), 51–53 (2007)
21. Song, W., Spencer, B., Du, W.: A transformation approach for classifying $\mathcal{ALCHI}(\mathcal{D})$ ontologies with a consequence-based $\mathcal{ALCH}$ reasoner. In: ORE. CEUR, vol. 1015, pp. 39–45 (2013)
22. Stefanoni, G., Motik, B., Horrocks, I.: Introducing nominals to the combined query answering approaches for EL. In: AAAI (2013)

23. Tsarkov, D., Horrocks, I.: FaCT++ Description Logic reasoner: System description. In: IJCAR. pp. 292–297 (2006)
24. Wu, Z., Eadon, G., Das, S., Chong, E.I., Kolovski, V., Annamalai, M., Srinivasan, J.: Implementing an inference engine for RDFS/OWL constructs and user-defined rules in Oracle. In: ICDE. pp. 1239–1248 (2008)
25. Zhou, Y., Cuenca Grau, B., Horrocks, I., Wu, Z., Banerjee, J.: Making the most of your triple store: query answering in OWL 2 using an RL reasoner. In: WWW (2013)